

# UC San Diego

## UC San Diego Previously Published Works

### Title

Pass-back chain extension expands multimodular assembly line biosynthesis.

### Permalink

<https://escholarship.org/uc/item/5kc033hc>

### Journal

Nature chemical biology, 16(1)

### ISSN

1552-4450

### Authors

Zhang, Jia Jia  
Tang, Xiaoyu  
Tang, Xiaoyu  
et al.

### Publication Date

2020

### DOI

10.1038/s41589-019-0385-4

Peer reviewed



Published in final edited form as:

*Nat Chem Biol.* 2020 January ; 16(1): 42–49. doi:10.1038/s41589-019-0385-4.

## Pass-back chain extension expands multi-modular assembly line biosynthesis

Jia Jia Zhang<sup>1</sup>, Xiaoyu Tang<sup>1</sup>, Tao Huan<sup>2</sup>, Avena C. Ross<sup>3</sup>, Bradley S. Moore<sup>1,4,\*</sup>

<sup>1</sup>Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California, USA.

<sup>2</sup>Department of Chemistry, University of British Columbia, Vancouver, Canada.

<sup>3</sup>Department of Chemistry, Queen's University, Kingston, Ontario, Canada.

<sup>4</sup>Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, California, USA.

### Abstract

Modular nonribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) enzymatic assembly lines are large and dynamic protein machines that generally effect a linear sequence of catalytic cycles. Here we report the heterologous reconstitution and comprehensive characterization of two hybrid NRPS-PKS assembly lines that defy many standard rules of assembly line biosynthesis to generate a large combinatorial library of cyclic lipodepsipeptide protease inhibitors called thalassospiramides. We generate a series of precise domain-inactivating mutations in thalassospiramide assembly lines and present evidence for an unprecedented biosynthetic model that invokes inter-module substrate activation and tailoring, module skipping, and pass-back chain extension, whereby the ability to pass the growing chain back to a preceding module is flexible and substrate-driven. Expanding bidirectional inter-module domain interactions could represent a viable mechanism for generating chemical diversity without increasing the size of biosynthetic assembly lines and challenges our understanding of the potential elasticity of multi-modular megaenzymes.

Modular nonribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) enzymes are molecular-scale assembly lines that construct complex polymeric products, many of which are useful to humans as life-saving drugs. The first characterized assembly lines exhibited an elegant co-linear biosynthetic logic, whereby the linear arrangement of functional units, called modules, along an NRPS or PKS polypeptide directly correlates to

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

\* bsmoore@ucsd.edu.

Author contributions

J.J.Z. and B.S.M. designed the study. J.J.Z., X.T., and A.C.R. performed cloning, gene deletion, and other molecular biology experiments. J.J.Z. and X.T. performed heterologous expression and chemical extraction experiments. J.J.Z. and T.H. performed mass spectrometry experiments and analyzed mass spectrometry data. J.J.Z. and B.S.M. analyzed all data and wrote the manuscript with input from all authors.

Competing financial interests

The authors declare no competing financial interests.

the chemical structure of the product<sup>1</sup>. The PKS giving rise to the antibiotic erythromycin<sup>2</sup> and the NRPS producing the antibiotic daptomycin<sup>3</sup> are two such examples. The core components of an NRPS or PKS assembly line elongation module include, respectively, condensation (C) or ketosynthase (KS) domains catalyzing chain extension, adenylation (A) or acyltransferase (AT) domains for substrate selection, and thiolation (T) domains for covalent substrate tethering. Optional tailoring domains such as methyltransferase (MT), ketoreductase (KR), or dehydratase (DH) domains, if present, chemically modify building blocks or chain-extension intermediates. This one-to-one correlation between product moieties and assembly line modules with the requisite catalytic domains is one feature that makes NRPS/PKS enzymes among the largest proteins found in nature.

However, it is now clear that many assembly lines do not strictly abide by the rules of co-linearity. A phylogenetically distinct class of modular PKSs are *trans*-AT PKSs, which do not directly encode AT domains within modules but instead as stand-alone enzymes that act *in trans*<sup>4</sup>. A separate type of NRPS, referred to as a nonlinear NRPS, deviates from the standard core domain arrangement of C-A-T and reuses a single domain more than once<sup>5</sup>. Several nonlinear NRPSs possess modules missing A domains and are presumably loaded by A domains from upstream modules<sup>6-10</sup>. Leveraging domain activities *in trans* or from different modules reduces the size of biosynthetic assembly lines and thus may represent a mechanism for minimizing modular assembly lines without sacrificing product complexity.

A particularly intriguing set of hybrid NRPS-PKS assembly lines that are both nonlinear and *trans*-AT are those responsible for the biosynthesis of the thalassospiramides, a large group of immunosuppressive cyclic lipodepsipeptides<sup>11-13</sup>. Thalassospiramide NRPS-PKS genes have been identified in several marine *Rhodospirillaceae* bacteria and exhibit distinct architectures that range in domain and module “completeness”<sup>13</sup>. While it is still unclear whether all configurations are functional, thalassospiramide assembly lines representing the most “complete” and “incomplete” architectures are both capable of producing a large combination of lipodepsipeptides that vary in fatty acid and amino acid composition, order, and length<sup>12</sup>. Previous work identifying these NRPS-PKS genes and structurally characterizing their numerous and diverse chemical products<sup>11-13</sup> led us to hypothesize that these assembly lines must operate with an unprecedented degree of nonlinearity. Furthermore, we posited that in order to generate their chemical products, thalassospiramide assembly lines must catalyze one or two rounds of pass-back chain extension, where the chain-extension intermediate is passed from a downstream module back to an upstream module within the same polypeptide.

Here, we report a comprehensive characterization of the thalassospiramide biosynthetic machinery from  $\alpha$ -proteobacteria *Thalassospira* sp. CNJ-328 and *Tistrella mobilis* KA081020-065, which represent the most “complete” and “incomplete” assembly line architectures, respectively<sup>13</sup>. We present an experimentally supported and mechanistically novel biosynthetic model that invokes inter-module substrate activation and tailoring, module skipping, and pass-back chain extension, whereby the ability to pass the growing chain forward or backward is flexible and influenced by the identity and chain length of the chemical intermediate. These newly described features accentuate the potential

bidirectionality and flexibility of multi-modular megaenzymes and reveal new engineering opportunities and structural considerations.

## Results

### Cloning and heterologous expression of *ttc* and *ttn*.

Close inspection of the thalassospiramide *ttc* and *ttn* gene clusters revealed that the NRPS/PKS genes from *Thalassospira* and *Tistrella*, respectively, occupy completely different genomic contexts (Fig. 1a,b, Supplementary Table 1). *Ttc* also includes a 4'-phosphopantetheinyl transferase (PPTase), TtcD, for which there is no homolog in *ttn*. Notably absent from the genomic vicinity of both pathways are any genes encoding stand-alone AT or A domains (Supplementary Table 1), although both assembly lines possess a *trans*-AT PKS module and Ttn contains two NRPS modules without A domains. We chose to clone a broader range for *ttc* and a more limited range for *ttn* (Fig. 1a,b).

To directly clone these gene clusters, a new bacterial artificial chromosome (BAC)-based transformation-associated recombination (TAR) cloning vector, pCAP-BAC (pCB), was designed and constructed to enable stable maintenance of large constructs in *Escherichia coli* (Supplementary Fig. 1). pCB lacks host-specific integration elements, which can be introduced after cloning, to make it easier to retrofit cloned pathways with integration elements for different hosts. Following successful cloning of *ttc*, heterologous expression was attempted but never achieved in *E. coli*, despite efforts to perform promoter refactoring, stabilize protein expression, and co-express the pathway with various promiscuous PPTases (Supplementary Fig. 2). Thus, we constructed a *Pseudomonas* integration cassette containing the Int-B13 site-specific recombinase<sup>14</sup> and introduced it into the vector backbone to generate pCB-*ttc*-int (Supplementary Fig. 1). This construct was successfully integrated into the genome of *Pseudomonas putida* EM383<sup>15</sup> (Supplementary Fig. 3). The same procedure was used for *ttn*, and expression of both gene clusters in *P. putida* was successful as evidenced by detection of the representative product thalassospiramide A (**1**) (Fig. 1c, Supplementary Fig. 4). This validated that the Ttc and Ttn assembly lines do not require additional pathway-specific enzymatic components, beyond what was transferred to the host and supplied through primary metabolism. To our knowledge, this is the first report of successful heterologous expression of a *trans*-AT pathway without co-transfer of a cognate AT. Our current hypothesis as to why expression was successful in *P. putida* but not in *E. coli* is that the *P. putida* primary metabolic AT is capable of interfacing with the *trans*-AT PKS modules, while the AT from *E. coli* is not.

### Reconstitution of thalassospiramide structural diversity.

We next explored whether the heterologously expressed Ttc and Ttn clusters could reproduce the full suite of thalassospiramide chemical diversity. Thalassospiramide cyclic lipodepsipeptides can be grouped into four categories based on their chemical structures and biosynthetic origin<sup>11-13</sup> (Fig. 2). Products of Ttc can incorporate serine, phenylalanine, or tyrosine as the first amino acid residue (colored blue in Fig. 2), which can be extended by a single ketide unit to generate “B-like” as opposed to “A1-like” thalassospiramides. Alternatively, Ttn incorporates serine or valine in the first position, then includes or omits a

valine residue (colored red in Fig. 2) to generate “A4-like” or “E-like” thalassospiramides, respectively. Both assembly lines produce “A-like” thalassospiramides at greater relative abundance than their “B-like” or “E-like” counterparts.

Additional elements that contribute to structural diversity include the N-terminal fatty acid, which is predominantly an atypical C10:1( 3) fatty acid, and the pattern of N-methylation, which is predominantly limited to the final tyrosine residue of the cyclic peptide core but can also extend to the adjacent valine residue for products of Ttc and further to valine within the linear peptide for Ttm (Fig. 2). Finally, the number of linear Ser-C2-Val units (4-amino-3,5-dihydroxy-N-pentanyl-valine, colored green in Fig. 2) can be 0, 1, or 2 for both assembly lines, which presumably arises from passage of chain-extension intermediates from module 4 back to module 2 or from module 5 back to module 3. Based on the different possible combinations of these variables, each assembly line can theoretically generate well over 100 compound analogs; several dozen compounds are routinely detected from small (50 mL) cultures of producing organisms using mass spectrometry.

LC-MS-MS analysis revealed that nearly all analogs detected from the native producers are also produced by the heterologous host, with some subtle differences in relative production levels (Supplementary Fig. 5). Overall titers of most analogs are comparable between native strains and host and in some cases greater in the host (Supplementary Fig. 5).

### Characterization of non-assembly line genes.

To determine whether co-transferred genes beyond the core NRPS/PKS genes affect thalassospiramide titer or product distribution, we performed targeted deletions of all non-assembly line genes within *ttc*. Previous work using the broad-host-range vector, pCAP05 (ref. <sup>16</sup>), demonstrated that no upstream genes (–7 through –1) are essential, although heterologous expression using this vector produced low yields (Supplementary Fig. 6a) and proved to be genetically unstable over time. Targeted deletion of *ttc* –1, +1, +2, +3, and +4 in the stable pCB-*ttc*-int expression construct had no impact on thalassospiramide production; however, deletion of the gene encoding the putative PPTase TtcD resulted in an approximately five-fold reduction in thalassospiramide A (**1**), which was restored upon genetic complementation of *ttcD* (Supplementary Figs. 6 and 7). This observation suggests that the single native *P. putida* PPTase is capable of activating Ttc carrier proteins, albeit not as effectively as TtcD. The *P. putida* PPTase is clearly capable of activating Ttm carrier proteins, as no cognate PPTase is encoded within *ttm*, and thalassospiramides are still produced in the heterologous host. However, co-expression of *ttcD* with *ttm* resulted in an approximately two-fold increase in levels of thalassospiramide A (**1**) (Supplementary Figs. 6 and 7). Quantitative analysis suggests that the PPTase TtcD favorably biases production of analogs that incorporate one or more linear Ser-C2-Val units ( $n \geq 1$ ), particularly for Ttm, as *ttcD* co-expression actually decreases production of C2 (**6**) and E4 (**25**) ( $n=0$ ) (Supplementary Fig. 7). This result suggests that TtcD-catalyzed phosphopantetheinylation of TtmA carrier proteins predisposes TtmA to perform pass-back chain extension at the expense of linear assembly through an unknown mechanism.

### Inactivation and testing of assembly line domains.

Taken together, the results of the heterologous expression and gene deletion experiments strongly suggest that thalassospiramide structural diversity is generated directly from the multi-modular assembly line itself and does not involve accessory enzymes beyond what is provided from primary metabolism. Although certain modules appear to be missing domains based on retro-biosynthetic analysis of thalassospiramide chemical structures, all necessary core and tailoring domains are present somewhere along the assembly line, including a DH domain in module 4 of both systems that was not previously annotated (Fig. 2, Supplementary Fig. 8). Thus, we set out to investigate the predicted inter-module activity of assembly line domains.

Our initial approach leveraged gene deletion and complementation tools, focusing first on the smaller *ttcC* that encodes the last four domains of terminal module 6. Despite it harboring the lone assembly line MT domain, some thalassospiramides, such as A8 (**5**), are unusual in containing a ‘misplaced’ penultimate N-methylated valine residue in addition to the conserved terminal N-methylated tyrosine. Although MTs are usually positioned at the C-terminus of interrupted A domains<sup>17</sup>, MT<sub>6</sub> is positioned at the N-terminus of A<sub>6</sub> (Supplementary Fig. 8d). To explore whether MT<sub>6</sub> can methylate amino acids activated by both the A<sub>5</sub> and A<sub>6</sub> adenylation domains, we deleted *ttcC*, resulting in complete loss of thalassospiramide production. Complementation with wild-type *ttcC* restored thalassospiramide production, while complementation with a mutant encoding TtcC-G234D, in which MT<sub>6</sub> has been selectively inactivated, resulted in dramatic reduction and complete loss of thalassospiramides A (**1**) and A8 (**5**), respectively, and concomitant formation of a new product with HR-MS and MS/MS spectra consistent with desmethyl thalassospiramide A, or thalassospiramide A15 (**26**) (Supplementary Table 2, Supplementary Fig. 9). The same result was observed for all analogs, resulting in the formation of a series of desmethyl cyclic lipodepsipeptides. This confirms our hypothesis that MT<sub>6</sub> can act within the upstream module 5 of TtcB. Furthermore, it suggests that pass-back chain extension occurs between modules 4 and 2 as opposed to 5 and 3, as promiscuous methylation is confined to the cyclic valine residue for Ttc. If linear valine residues were installed by module 5, promiscuous methylation should extend to these positions; however, we never observe this for products of Ttc. As module 5 of Ttm does not possess an A domain and presumably borrows the activity of A<sub>2</sub>, this could explain why promiscuous methylation can extend to upstream valine residues for Ttm but not Ttc.

We attempted to use the same approach to characterize other thalassospiramide domains; however, efforts to perform PCR mutagenesis in *ttcA*, which is over 6.5 kb, and *ttcB* and *ttnA*, which are both over 15.5 kb, were ultimately unsuccessful. Therefore, we established new methodology combining oligo recombineering with CRISPR-Cas9 counter-selection for facile introduction of point mutations to large DNA constructs cloned into the pCB vector backbone (Supplementary Fig. 10). Using this method, we selectively inactivated a series of assembly line domains (Supplementary Fig. 11) to directly interrogate their role in thalassospiramide biosynthesis.

Consistent with our annotation of C<sub>1a</sub> of TtcA as a starter C domain responsible for fatty acylation of the first amino acid residue<sup>18</sup>, C<sub>1a</sub> inactivation completely abolished production of all thalassospiramides (Fig. 3a). We did not detect any masses corresponding to core peptides lacking an N-terminal fatty acid, suggesting that the assembly line can only generate lipopeptide products. In contrast, inactivation of A<sub>1a</sub> using two different point mutations, TtcA-G631D and TtcA-K972A, resulted in essentially complete loss of all thalassospiramides incorporating phenylalanine or tyrosine as the first residue but maintained or enhanced production of analogs incorporating serine in the first position (Fig. 3a). Inactivation of A<sub>3</sub> using the analogous lysine to alanine mutation (TtcB-K2045A) resulted in complete loss of thalassospiramide production, as did selective inactivation of T<sub>1a</sub> (Fig. 3a). These results suggest that the serine residue adenylated by A<sub>3</sub> is directly loaded onto T<sub>1a</sub> during biosynthesis of analogs that incorporate serine as the first amino acid residue.

We hypothesized that “B-like” thalassospiramides from Ttc arise through PKS module 1b and, correspondingly, that “A1-like” thalassospiramides arise through module 1b skipping. Consistent with that hypothesis, mutation of the active site cysteine of KS<sub>1b</sub> to alanine resulted in complete loss of “B-like” analogs but maintained production of “A1-like” analogs, although some yields were slightly reduced (Fig. 3a). Surprisingly, although AT<sub>1b</sub> inactivation (TtcA-S1728A) reduced production of “B-like” analogs, almost all could still be detected by LC-MS at ~8-24% of wild-type production levels (Fig. 3a). As the AT<sub>1b</sub> mutation is expected to abolish enzymatic activity and the assembly line has no other AT domains, we propose that the module 4-interacting *trans*-AT can partially complement AT<sub>1b</sub>.

We previously hypothesized that the tandem T domains in module 4 might be important for determining whether chain-extension intermediates are passed forward or backward<sup>12</sup>, although literature precedents indicated that multiple T domains do not change product identity but instead increase flux or yield<sup>19</sup>. We attempted to use the same oligo recombineering/CRISPR-Cas9 method to selectively inactivate T<sub>4a</sub> and T<sub>4b</sub>. However, CRISPR-Cas9 targeting did not result in oligo incorporation but instead recombination across the two very similar T domain sequences to generate an in-frame deletion of residues 3511 to 3596 in TtcB. The resultant mutant protein contained only a single chimeric T domain composed of 46% of the N-terminus of T<sub>4a</sub> and 54% of the C-terminus of T<sub>4b</sub> (Supplementary Fig. 11). Transfer of this construct to the heterologous host revealed that all thalassospiramide analogs could still be produced, including those incorporating linear Ser-C2-Val units arising from pass-back chain extension, albeit in decreased yields (Fig. 3a). This result proves that tandem T domains are not essential for bi-directional chain extension. Moreover, production of all analogs was affected equally, supporting the flux hypothesis.

Finally, we predicted that “E-like” thalassospiramides produced by Ttm arise from module 2 skipping, analogous to module 1b skipping in Ttc. Thus, if pass-back chain extension occurs through modules 5 and 3, C<sub>2</sub> inactivation in Ttm would preserve production of all “E-like” analogs, as C<sub>2</sub> is completely skipped in this model. While C<sub>1</sub> inactivation abolished production of all thalassospiramides, C<sub>2</sub> inactivation dramatically reduced and completely abolished production of thalassospiramides E (**23**) and E1 (**24**), respectively (Fig. 3b). This result confirms our hypothesis that most pass-back events occur between modules 4 and 2,



as levels of “E-like” analogs were not preserved. Furthermore, we observed a greater than 100-fold increase in levels of thalassospiramide E4 (**25**), suggesting that the inability to pass growing chains back via C<sub>2</sub> forces the assembly line to pass intermediates forward, resulting in enhanced production of the “premature” termination product E4 (**25**) (Fig. 3b). However, we also observed the formation of two new compounds not previously detected with HR-MS and MS/MS spectra consistent with the structures of thalassospiramides E5 (**27**) and E6 (**28**), indicating that pass-back between modules 5 and 3 can occur but correlates with additional substrate dehydration by module 4 (Fig. 3c, Supplementary Table 2).

### Models for thalassospiramide biosynthesis.

The results of Ttm C<sub>2</sub> inactivation provide a clear mechanistic insight into the flexibility and control of pass-back chain extension during thalassospiramide biosynthesis, as illustrated in Figure 4. Selective inactivation of C<sub>2</sub> does not affect early stages of “E-like” thalassospiramide biosynthesis, during which A<sub>2</sub> loads T<sub>1</sub> with valine and then module 2 is skipped following appendage of the N-terminal fatty acid. Assembly then proceeds through modules 3 and 4, but the chain-extension intermediate does not undergo immediate dehydration and may not be directly accessible to DH<sub>4</sub>. Now, the assembly line would normally pass the chain-extension intermediate from module 4 back to module 2 via C<sub>2</sub>, which is favored based on product distribution, as thalassospiramide E4 (**25**) is normally produced at very low abundance. We propose that the assembly line C domains play an important role in “measuring” intermediate chain length, promoting donation of the module 4 intermediate backward to module 2 instead of forward to module 5. However, C<sub>2</sub> inactivation forces the chain-extension intermediate forward, making it accessible to DH<sub>4</sub> before it enters module 5. Once within module 5, the intermediate would normally proceed directly to module 6, resulting in formation of thalassospiramide E4 (**25**). Consistent with this proposal, C<sub>2</sub> inactivation drives a substantial increase in levels of E4 (**25**) compared to wild-type. However, chain length can also be “measured” at the donor site of C<sub>6</sub> (as C<sub>6</sub> usually accepts donor substrates of longer chain length), prompting the assembly line to catalyze pass-back of a subset of intermediates that have already undergone dehydration from module 5 back to module 3, resulting in eventual formation of new products E5 (**27**) and E6 (**28**). This result provides additional evidence that the substrate becomes accessible to DH<sub>4</sub> just as it is passed forward to module 5 and dehydration is a passive result of forward chain extension, since a model that invokes DH<sub>4</sub> gating of module 5 entry would not be consistent with the observed increase in levels of E4 (**25**) or formation of E5 (**27**) and E6 (**28**). Furthermore, it suggests that biosynthesis is flexible and controlled by a mechanism that measures intermediate chain length. If we assume that dehydration of the hydroxyl group within the linear unit is a signature for module 5 progression, there is evidence that pass-back between modules 5 and 3 occurs at low frequency under normal conditions, perhaps as an additional checkpoint, as we can detect thalassospiramides E7 (**29**), E8 (**30**), and E9 (**31**) by LC-MS (Supplementary Table 2), which are all produced by wild-type Ttm and have undergone additional rounds of “premature” dehydration.

We can also propose a full model for thalassospiramide A (**1**) biosynthesis via Ttc (Fig. 5). T<sub>1a</sub> is preferentially adenylated with serine through the downstream A<sub>3</sub> domain. To our knowledge, this is the first report of an A domain activating a carrier protein within an



upstream module that already possesses its own active A domain<sup>12</sup>. Subsequently, module 1b is skipped during formation of “A1-like” analogs. Normal chain extension proceeds from modules 2 to 4, at which point we hypothesize that the substrate is sequestered from DH<sub>4</sub> activity and module 5 entry based on chain length. We propose that during the formation of “B-like” thalassospiramides, ketoreduction to generate the upstream statine-like amino acid residue occurs at this time, analogous to dual ketoreduction of two disparate positions catalyzed by KR<sub>3</sub> of PksJ during bacillaene biosynthesis<sup>20</sup>. The chain-extension intermediate within module 4 is then passed back to module 2 via C<sub>2</sub>, where it undergoes another round of linear chain extension through modules 2 to 4 (Fig. 5). Now, the intermediate has reached “sufficient” chain length and becomes accessible to DH<sub>4</sub> before transfer to module 5 via C<sub>5</sub>. MT<sub>6</sub> promiscuously methylates valine residues installed by module 5 to generate analogs such as thalassospiramide A8 (**5**). Finally, normal progression through modules 5 and 6 results in the formation of thalassospiramide A (**1**), the most abundant product of both Ttc and Ttm.

## Discussion

Modular assembly lines are large and dynamic enzymes that undergo dramatic domain conformational rearrangements during a single catalytic cycle<sup>21-24</sup>. However, whether multi-modular enzymes adopt a rigid, supermodular architecture<sup>25</sup> or a more flexible configuration<sup>26</sup> remains in debate. In this work, we demonstrate the ability of multi-modular assembly lines to catalyze bidirectional and nonlinear passage of chain-extension intermediates, favoring a more flexible arrangement.

Several new features of multi-modular NRPS/PKS biosynthesis are described in this work. Thalassospiramide assembly lines catalyze several instances of inter-module substrate activation and tailoring. While A domain supplementation has been previously reported<sup>6-10</sup>, prior examples have been limited to upstream domains supplementing downstream modules, often encoded on separate proteins, and only within modules that lack their own A domains. For Ttc, A<sub>3</sub> adenylates T<sub>1a</sub> with serine more frequently than A<sub>1a</sub> does with phenylalanine or tyrosine, although A<sub>1a</sub> is active and participates in biosynthesis of numerous thalassospiramide analogs. For Ttm, both A<sub>2</sub> and A<sub>3</sub> deliver substrates to T<sub>1</sub>, but only A<sub>3</sub> adenylates T<sub>5</sub>. Furthermore, only A<sub>2</sub> loads T<sub>1</sub> if module 2 is skipped, as all “E-like” analogs incorporate valine as the first amino acid, perhaps due to the specificity of C<sub>3</sub><sup>27</sup>. MT<sub>6</sub> promiscuously methylates valine residues activated by A<sub>5</sub> of Ttc and A<sub>2</sub> of Ttm. Finally, the *trans*-AT interacting with module 4 can partially supplement lost AT<sub>1b</sub> activity in Ttc. Complementation of lost *cis*-AT activity with *cis*-ATs from other modules or non-cognate *trans*-ATs has been previously observed in the 6-deoxyerythronolide B PKS synthase<sup>28</sup>.

Thalassospiramide assembly lines also catalyze programmed module skipping. Forward module skipping has been previously observed in PKS<sup>29-31</sup> and NRPS<sup>32-34</sup> systems, both naturally and as a byproduct of engineering. Ttc and Ttm catalyze analogous skipping of PKS module 1b and NRPS module 2, respectively, although NRPS module 2 does not sit at an enzyme junction but is embedded within a very large polypeptide. Perhaps as a result, skipping is favored in Ttc but disfavored in Ttm, based on product distribution.

Finally, thalassospiramide assembly lines catalyze pass-back chain extension. To our knowledge, this mechanism has not been previously reported in modular assembly line systems, although it is analogous to “iteration” observed in the fungal beauvericin and bassianolide synthetases<sup>35</sup>, where intermediates are passed back and forth between adjacent modules. Consistent with previous findings, tandem T domains in thalassospiramide assembly lines increase flux<sup>19</sup> but are not mechanistic determinants of pass-back chain extension. This result is also consistent with the observation that the thalassospiramide assembly line from *Oceanibaculum pacificum* contains only a single T domain in module 4 and can still produce thalassospiramide A (**1**)<sup>13</sup>. There is no evidence that additional T domains promote “stalling” to allow for additional tailoring reactions, as having a single T domain in module 4 does not change product distribution but decreased levels of all analogs equally.

Under normal conditions, thalassospiramide intermediates are passed from module 4 back to module 2 via C<sub>2</sub>. Upon C<sub>2</sub> inactivation, we observe pass-back chain extension from module 5 back to module 3 via C<sub>3</sub>. This suggests that the assembly line is flexible and can respond to perturbation, trying to “correct” for aberrant intermediate chain length caused by C<sub>2</sub> inactivation by passing back through C<sub>3</sub>, which, like C<sub>6</sub>, is specific for accepting intermediates with C-terminal valine residues. Our results are consistent with recent findings that multi-modular NRPSs can be “mixed and matched” if the specificity and relative position of downstream C domains are maintained<sup>27</sup>. It also suggests that assembly lines must possess some symmetry for pass-back chain extension to occur. While all NRPS modules possess mechanisms to control the timing of chain extension to prevent misinitiation<sup>36,37</sup>, thalassospiramide modules have the added ability to accept longer intermediates from nonsequential modules.

Perhaps more importantly, how physical proximity between upstream C domains and downstream T domains is achieved during pass-back chain extension remains unclear. We cannot predict the plasticity of the hybrid thalassospiramide assembly line proteins based on their primary amino acid sequence alone (Supplementary Fig. 8). Although we can assume that TtcA, TtcB, and TtmA are homodimeric due to the dimeric nature of KS and many linker domains, we do not know whether oligomerization impacts thalassospiramide assembly, although it is tempting to suggest that higher-order architecture helps facilitate nonlinear transfer.

It is curious that although the early stages of thalassospiramide biosynthesis are flexible, resulting in production of lipopeptides with a high degree of N-terminal structural diversity, the final stages are rather fixed. The C-terminal cyclic depsipeptide core of all thalassospiramide analogs is highly conserved, particularly the 12-membered ring and the  $\alpha,\beta$ -unsaturated carbonyl moiety that together form the pharmacophore responsible for calpain protease inhibition<sup>38</sup>. Thus, the assembly line constructs a series of chemical products in which the structural elements that confer activity are maintained, while accessory elements such as the fatty acid and linear chain composition and length, which may confer target specificity, are variable. It has been speculated that biosynthetic promiscuity resulting in chemical diversity may be evolutionarily advantageous<sup>39,40</sup>. Furthermore, the type of combinatorial biosynthesis observed in this study expands the

portfolio of small molecules produced without introducing new assembly line modules or domains, or even additional tailoring enzymes, thus representing a means of expanding chemical diversity while minimizing genomic space.

Our work was facilitated by new methodology combining oligo recombineering with CRISPR-Cas9 counter selection for facile editing of large DNA constructs. Although our method was used solely for domain inactivation in this study, it can be easily applied to perform other forms of multi-modular assembly line engineering, for example to alter A domain specificity<sup>41,42</sup>. Future efforts to understand the specific elements that enable and control assembly line flexibility will hopefully enhance efforts to engineer NRPS/PKS proteins and possibly expand their biosynthetic repertoire.

## Methods

### General methods.

A complete list of the primers, plasmids, and strains used in this study can be found in Supplementary Table 3. DNA fragments larger than 3 kb were amplified with PrimeSTAR Max (Clontech Laboratories, Inc.); all other PCR products were amplified with PrimeSTAR HS DNA polymerase (Clontech Laboratories, Inc.). DNA isolations and manipulations were carried out using standard protocols. *Thalassospira* sp. CNJ-328 and *T. mobilis* KA081020-065 were grown in GYP media (glucose 10 g/L, yeast extract 4 g/L, peptone 2 g/L, sea salt 25 g/L). *S. cerevisiae* VL6-48N was grown in YPDA media (yeast extract 10 g/L, peptone 20 g/L, dextrose 20 g/L, adenine 100 mg/L) or selective histidine drop-out media containing 5-FOA (yeast nitrogen base without amino acids and ammonium sulfate 1.7 g/L, yeast synthetic dropout medium without histidine 1.9 g/L, sorbitol 182 g/L, dextrose 20 g/L, ammonium sulfate 5 g/L, adenine 100 mg/L, 5-FOA 1 g/L). *E. coli* and *P. putida* strains were grown in LB. *E. coli* TOP10 and DH5 $\alpha$   $\lambda$ pir were used for standard cloning procedures. *E. coli* BW25113/pIJ790 was used for  $\lambda$  Red PCR targeting, and *E. coli* HME68 was used for oligo recombineering and CRISPR-Cas9 counter selection. *P. putida* EM383 was used for heterologous expression. All strains were grown at 30 °C except TOP10 and DH5 $\alpha$   $\lambda$ pir, which were grown at 37 °C. Liquid cultures were grown shaking at 220 r.p.m. When necessary, *E. coli* (and *P. putida*) cultures were supplemented with the following antibiotics: 50  $\mu$ g/mL kanamycin (150  $\mu$ g/mL for *P. putida*), 10  $\mu$ g/mL gentamycin (30  $\mu$ g/ml for *P. putida*), 50  $\mu$ g/mL apramycin, 100  $\mu$ g/mL ampicillin, 25  $\mu$ g/mL chloramphenicol.

### Cloning and heterologous expression of *ttc* and *ttm*.

The overall workflow for genetic manipulation and heterologous expression of the *ttc* and *ttm* pathways is outlined in Supplementary Fig. 1. Biosynthetic gene clusters were cloned from genomic DNA using a TAR cloning protocol described previously<sup>43</sup>. Cluster-specific capture vectors were generated through a one-step PCR amplification of pCAP-BAC (pCB) using primers pCB-*ttc*CV\_F/R for *ttc* and pCB-*ttm*CV\_F/R for *ttm*. Yeast clones were screened, and PCR positive constructs were purified and transferred to *E. coli* TOP10 for verification by restriction digestion. pCB was miniprepmed from at least 25 mL of VL6-48N and at least 10 mL of TOP10. Expression was attempted but never achieved in *E. coli* strains, including BL21(DE3) and BAP1<sup>44</sup> (Supplementary Fig. 2). pJZ001, containing the

*intB13* cassette, was assembled by Gibson Assembly using four PCR fragments (amplified using primers CEN6/ARS4\_608F/R, *intB13\_2330F/R*, *aacC1\_1257F/R*, and *pADH\_597F/R*) and the pACYCDuet-1 vector backbone, linearized using HindIII and XhoI. Fully assembled pJZ001 was digested using HindIII and XhoI, and the 5155 bp fragment was gel purified and used for  $\lambda$  Red recombination<sup>45</sup> to knock-in the *intB13* cassette into the pCB vector backbone, replacing several yeast genes no longer necessary. Retrofitted constructs were then transferred to *P. putida* by electroporation as described previously<sup>16</sup> and selected for using kanamycin and gentamicin. *IntB13*-mediated integration of *ttc* into the genome of *P. putida* was characterized by PCR and chemical analysis (Supplementary Fig. 3). Edited constructs were similarly introduced to *P. putida* and then tested for heterologous production of lipopeptide products.

### Extraction and LC-MS analysis.

Precultures were inoculated with colonies picked from plates and grown overnight before being inoculated into full 50 mL cultures in 250 mL Erlenmeyer flasks (in triplicate). Full cultures were grown for 5 hours before addition of 1.5 g of autoclaved XAD7HP resin per 50 mL of culture. After 24 hours of additional incubation, culture ODs were measured and recorded at 600 nm and supernatant and cells were decanted. Resin was washed three times with Milli-Q water before extraction with 20 mL of ethyl acetate. Extracts were dried under nitrogen, resuspended in 200  $\mu$ L of methanol, and filtered through a 0.22  $\mu$ m filter prior to LC-MS-MS analysis.

An Agilent 1100 series HPLC system (Palo Alto, CA, U.S.A.) was coupled to a Bruker Impact II Q-TOF mass spectrometer (Billerica, MA, U.S.A.) for LC-MS analysis. An Agilent ZORBAX 300SB-C18 LC column (300  $\text{\AA}$ , 5  $\mu$ m, 150  $\times$  0.5 mm) was used for LC separation. Mobile phase A was H<sub>2</sub>O in 0.1% FA and mobile phase B was ACN in 0.1% FA. The LC gradient was: t=0.00 min, 70% A; t=3.00 min, 70% A; t=5.00 min, 57% A; t=35.00 min, 57% A; t=49.00 min, 20% A; t=51.00 min, 0% A; t=52.00 min, 0% A; t=57.00 min, 70% A; t=60.00 min, 70% A. A post time of 10 min was set to re-equilibrate the column. For shorter runs, the LC gradient was: t=0.00 min, 70% A; t=3.00 min, 70% A; t=23.00 min, 20% A; t=24.00 min, 0% A; t=27.00 min, 70% A; t=30.00 min, 70% A. A post time of 3 min was set to re-equilibrate the column. Flow rate was 20  $\mu$ L/min. Sample injection volume was 2  $\mu$ L.

MS conditions for MS/MS spectra generation were set as follows: capillary voltage, 4500; nebulizer gas flow, 0.8 Bar; dry gas, 5.0 L/min at 180  $^{\circ}$ C; funnel 1 RF 150 Vpp; funnel 2 RF, 300 Vpp; isCID energy, 0 eV; hexapole RF: 50 Vpp; Quadrupole ion energy, 4 eV; low mass 50 m/z; collision cell energy, 20 – 50 eV; pre pulse storage 5.0  $\mu$ s; collision RF, ramp from 350 to 800 Vpp; transfer time ramp from 50 to 100  $\mu$ s; detection mass range 25 to 1000 m/z; MS/MS spectra collection rate was 2.0 Hz.

All samples analyzed by comparison were run at the same time and under the same conditions. Values were normalized by culture ODs and compared only for peaks with identical MS spectra and retention time. HR-MS data for thalassospiramide analogs analyzed in this study are provided in Supplementary Table 2 and Supplementary Fig. 12-24.

### Gene deletion and complementation experiments.

Gene deletions were made using  $\lambda$  Red PCR targeting as described previously<sup>46</sup>. Primers used to amplify the *aac(3)IV* cassette and confirm gene deletions are listed in Supplementary Table 3. Deletions using this cassette were made after addition of the *intB13* cassette, which contains a gentamycin resistance gene, as the apramycin resistance gene *aac(3)IV* confers resistance to gentamycin in *E. coli*. For complementation, *ttcD* and *ttcC* were amplified using primers Tn7-ttcD\_F/R and Tn7-ttcC\_F/R, respectively, and cloned into the mini-Tn7 vector pUC18R6K-mini-Tn7T-Gm<sup>47</sup>. Cloned constructs were introduced to *P. putida* by electroporation<sup>16</sup> along with the helper plasmid pTNS1<sup>47</sup> and selected for using gentamycin. Complemented *P. putida* clones were then made electrocompetent and pCB constructs were transferred by electroporation and selected for using kanamycin and gentamycin. For complementation of TtcC-G234D (MT<sub>6</sub> inactivation), the mutation was generated by amplification of pTn7::*ttcC* using primers Tn7-ttcC-g702a\_F/R and confirmed by sequencing. Although a *ttcA* deletion construct was generated and a mini-Tn7 vector containing *ttcA* was prepared, the latter ultimately could not be transferred to *P. putida* for complementation, as no clones were obtained even after multiple attempts, likely due to the large size of the gene (>6.5 kb).

### Inactivation and testing of assembly line enzymatic domains.

Assembly line domain active sites are shown in Supplementary Fig. 8. Motifs and active sites were identified by sequence alignment against annotated NRPS/PKS domains<sup>22-24,48-55</sup>. pJZ002, an ampicillin resistant version of the pCas9<sup>56</sup> vector, was constructed as follows. First, the BsaI restriction site was first removed from the ampicillin resistance gene *bla* via PCR amplification of pKD20 using primers pKD20-g848a\_F/R. The resulting construct was PCR amplified using primers ts-repA101\_F/R and combined with a fragment amplified from pCas9 using primers pCas9\_5058F/R by Gibson assembly. Spacer sequences were cloned into pJZ002 as described previously<sup>56</sup>. Spacer sequences and targeting oligos used to target specific domains are listed in Supplementary Table 3. Oligo recombination and CRISPR-Cas9 counter selection were performed as described previously<sup>56,57</sup>, with several modifications. The general workflow is shown in Supplementary Fig. 10. pCB constructs were first transferred to *E. coli* HME68 by electroporation and selected for using kanamycin. A transformant was picked and grown at 30 °C to OD<sub>600</sub> 0.4-0.5 and heat shocked for 15 minutes at 42 °C in a shaking water bath before being chilled on ice for 10 minutes. The cells were pelleted and washed with ice-cold water before being resuspended in a small volume of ice-cold water. 100 ng of pJZ002 containing the appropriate spacer was mixed with 100 ng of targeting oligo and the DNA mixture was introduced to the cells prior to electroporation at 2.5 kV in a 2 mm gap electroporation cuvette. Cells were recovered for 2 hours at 30 °C shaking and plated on LB with kanamycin and ampicillin. Four clones of each mutant were picked, minipreped, and screened by sequencing. If no correct mutant was identified, a new spacer sequence was designed and cloned into pJZ002 and the method was retried. Very subtle mutations could be recovered efficiently using an effective spacer sequence, although the effectiveness of the spacer could only be determined empirically. In total, four of 12 spacer sequences were redesigned to achieve successful editing (Supplementary Table 3). Correctly edited

constructs were transferred to TOP10 and confirmed by restriction digestion (Supplementary Fig. 11) before being transferred to *P. putida* for heterologous expression.

### Data availability

The *ttc* and *ttm* biosynthetic gene cluster sequences are available in the MIBiG database (accession BGC0001050 and BGC0001876). Plasmids pCAP-BAC (#120229), pJZ001 (#120230), and pJZ002 (#120231) are available at Addgene.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgements

The authors thank J.R. van der Meer (University of Lausanne) for providing plasmid pRMR6K-Gm, V. de Lorenzo (National Center for Biotechnology-CSIC) for providing strain *P. putida* EM383, and D.L. Court (National Cancer Institute, NIH) for providing strain *E. coli* HME68. We are grateful to A. Edlund (J. Craig Venter Institute), P.Y. Qian (Hong Kong University of Science and Technology), H. Xia (Shanghai Institutes for Biological Sciences, CAS), and W. Fenical and P.R. Jensen (Scripps Institution of Oceanography, UCSD) for facilitating access to equipment, chemical standards, and bacterial strains. We also thank Y. Kudo, P.A. Jordan, J.R. Chekan, L.T. Hoang, and S. Carreto for helpful discussion and technical assistance. The research reported has been supported by National Institutes of Health grants F31-AI129299 to J.J.Z. and R01-GM085770 to B.S.M.

### References

1. Weissman KJ The structural biology of biosynthetic megaenzymes. *Nat. Chem. Biol.* 11, 660–670, doi:10.1038/nchembio.1883 (2015). [PubMed: 26284673]
2. Cane DE Programming of erythromycin biosynthesis by a modular polyketide synthase. *J. Biol. Chem.* 285, 27517–27523, doi:10.1074/jbc.R110.144618 (2010). [PubMed: 20522551]
3. Robbel L & Marahiel MA Daptomycin, a bacterial lipopeptide synthesized by a nonribosomal machinery. *J. Biol. Chem.* 285, 27501–27508, doi:10.1074/jbc.R110.128181 (2010). [PubMed: 20522545]
4. Helfrich EJ & Piel J Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* 33, 231–316, doi:10.1039/c5np00125k (2016). [PubMed: 26689670]
5. Sussmuth RD & Mainz A Nonribosomal peptide synthesis-principles and prospects. *Angew. Chem. Int. Ed. Engl.* 56, 3770–3821, doi:10.1002/anie.201609079 (2017). [PubMed: 28323366]
6. Magarvey NA, Haltli B, He M, Greenstein M & Hucul JA Biosynthetic pathway for mannopeptimycins, lipoglycopeptide antibiotics active against drug-resistant gram-positive pathogens. *Antimicrob. Agents Chemother.* 50, 2167–2177, doi:10.1128/AAC.01545-05 (2006). [PubMed: 16723579]
7. Felnagle EA, Rondon MR, Berti AD, Crosby HA & Thomas MG Identification of the biosynthetic gene cluster and an additional gene for resistance to the antituberculosis drug capreomycin. *Appl. Environ. Microbiol.* 73, 4162–4170, doi:10.1128/AEM.00485-07 (2007). [PubMed: 17496129]
8. Thomas MG, Chan YA & Ozanick SG Deciphering tuberactinomycin biosynthesis: isolation, sequencing, and annotation of the viomycin biosynthetic gene cluster. *Antimicrob. Agents Chemother.* 47, 2823–2830 (2003). [PubMed: 12936980]
9. Du L, Sanchez C, Chen M, Edwards DJ & Shen B The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. *Chem. Biol.* 7, 623–642 (2000). [PubMed: 11048953]
10. Gehring AM et al. Iron acquisition in plague: modular logic in enzymatic biogenesis of yersiniabactin by *Yersinia pestis*. *Chem. Biol.* 5, 573–586 (1998). [PubMed: 9818149]

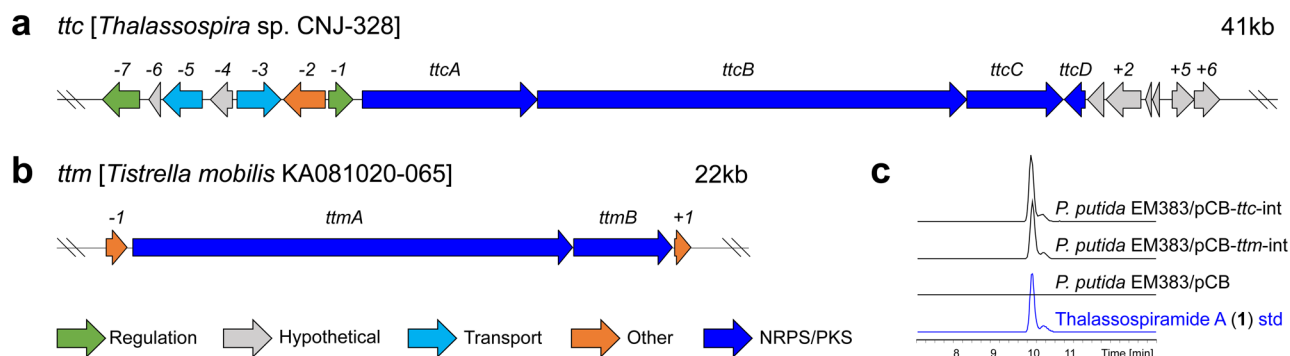


11. Oh DC, Strangman WK, Kauffman CA, Jensen PR & Fenical W Thalassospiramides A and B, immunosuppressive peptides from the marine bacterium *Thalassospira* sp. Org. Lett. 9, 1525–1528, doi:10.1021/ol070294u (2007). [PubMed: 17373804]
12. Ross AC et al. Biosynthetic multitasking facilitates thalassospiramide structural diversity in marine bacteria. J. Am. Chem. Soc. 135, 1155–1162, doi:10.1021/ja3119674 (2013). [PubMed: 23270364]
13. Zhang W et al. Family-wide structural characterization and genomic comparisons decode the diversity-oriented biosynthesis of thalassospiramides by marine Proteobacteria. J. Biol. Chem 291, 27228–27238, doi:10.1074/jbc.M116.756858 (2016). [PubMed: 27875306]
14. Miyazaki R & van der Meer JR A new large-DNA-fragment delivery system based on integrase activity from an integrative and conjugative element. Appl. Environ. Microbiol. 79, 4440–4447, doi:10.1128/AEM.00711-13 (2013). [PubMed: 23686268]
15. Martinez-Garcia E, Nikel PI, Aparicio T & de Lorenzo V *Pseudomonas* 2.0: genetic upgrading of *P. putida* KT2440 as an enhanced host for heterologous gene expression. Microb. Cell Fact. 13, 159, doi:10.1186/s12934-014-0159-3 (2014). [PubMed: 25384394]
16. Zhang JJ, Tang X, Zhang M, Nguyen D & Moore BS Broad-host-range expression reveals native and host regulatory elements that influence heterologous antibiotic production in Gram-negative bacteria. MBio 8, e01291, doi:10.1128/mBio.01291-17 (2017). [PubMed: 28874475]
17. Labby KJ, Watsula SG & Garneau-Tsodikova S Interrupted adenylation domains: unique bifunctional enzymes involved in nonribosomal peptide biosynthesis. Nat. Prod. Rep. 32, 641–653, doi:10.1039/c4np00120f (2015). [PubMed: 25622971]
18. Rausch C, Hoof I, Weber T, Wohlleben W & Huson DH Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. BMC Evol. Biol. 7, 78, doi:10.1186/1471-2148-7-78 (2007). [PubMed: 17506888]
19. Crosby J & Crump MP The structural role of the carrier protein--active controller or passive carrier. Nat. Prod. Rep. 29, 1111–1137, doi:10.1039/c2np20062g (2012). [PubMed: 22930263]
20. Calderone CT, Bumpus SB, Kelleher NL, Walsh CT & Magarvey NA A ketoreductase domain in the PksJ protein of the bacillaene assembly line carries out both alpha- and beta-ketone reduction during chain growth. Proc. Natl. Acad. Sci. U S A 105, 12809–12814, doi:10.1073/pnas.0806305105 (2008). [PubMed: 18723688]
21. Whicher JR et al. Structural rearrangements of a polyketide synthase module during its catalytic cycle. Nature 510, 560–564, doi:10.1038/nature13409 (2014). [PubMed: 24965656]
22. Drake EJ et al. Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. Nature 529, 235–238, doi:10.1038/nature16163 (2016). [PubMed: 26762461]
23. Reimer JM, Aloise MN, Harrison PM & Schmeing TM Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. Nature 529, 239–242, doi:10.1038/nature16503 (2016). [PubMed: 26762462]
24. Miller BR, Drake EJ, Shi C, Aldrich CC & Gulick AM Structures of a Nonribosomal Peptide Synthetase Module Bound to MbtH-like Proteins Support a Highly Dynamic Domain Architecture. J. Biol. Chem. 291, 22559–22571, doi:10.1074/jbc.M116.746297 (2016). [PubMed: 27597544]
25. Marahiel MA A structural model for multimodular NRPS assembly lines. Nat. Prod. Rep. 33, 136–140, doi:10.1039/c5np00082c (2016). [PubMed: 26429504]
26. Tarry MJ, Haque AS, Bui KH & Schmeing TM X-ray crystallography and electron microscopy of cross- and multi-module nonribosomal peptide synthetase proteins reveal a flexible architecture. Structure 25, 783–793 e784, doi:10.1016/j.str.2017.03.014 (2017). [PubMed: 28434915]
27. Bozhuyuk KAJ et al. *De novo* design and engineering of non-ribosomal peptide synthetases. Nat. Chem. 10, 275–281, doi:10.1038/nchem.2890 (2018). [PubMed: 29461518]
28. Dunn BJ, Watts KR, Robbins T, Cane DE & Khosla C Comparative analysis of the substrate specificity of trans- versus cis-acyltransferases of assembly line polyketide synthases. Biochemistry 53, 3796–3806, doi:10.1021/bi5004316 (2014). [PubMed: 24871074]
29. Awakawa T et al. Salinipyrone and pacificanone are biosynthetic by-products of the rosamicin polyketide synthase. Chembiochem 16, 1443–1447, doi:10.1002/cbic.201500177 (2015). [PubMed: 25930739]



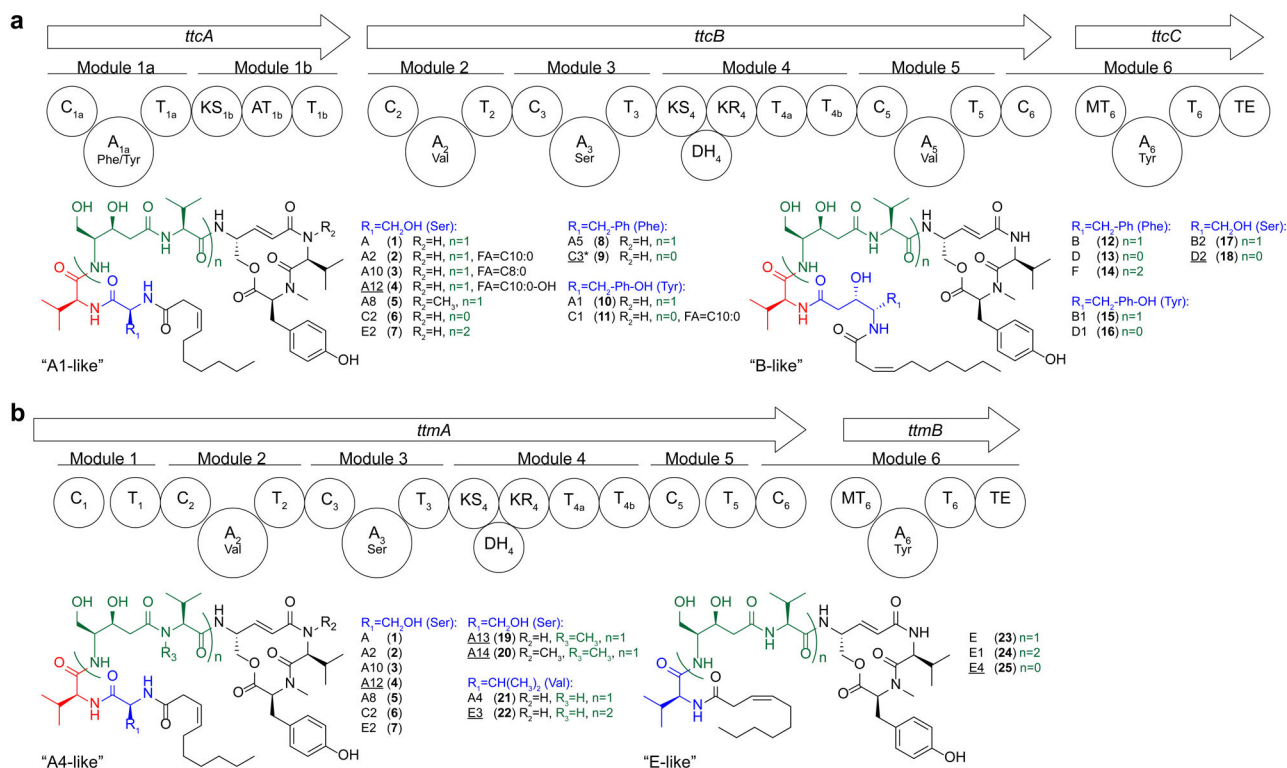
30. Moss SJ, Martin CJ & Wilkinson B Loss of co-linearity by modular polyketide synthases: a mechanism for the evolution of chemical diversity. *Nat. Prod. Rep.* 21, 575–593, doi:10.1039/b315020h (2004). [PubMed: 15459756]
31. Thomas I, Martin CJ, Wilkinson CJ, Staunton J & Leadlay PF Skipping in a hybrid polyketide synthase. Evidence for ACP-to-ACP chain transfer. *Chem. Biol.* 9, 781–787 (2002). [PubMed: 12144921]
32. Wenzel SC, Meiser P, Binz TM, Mahmud T & Muller R Nonribosomal peptide biosynthesis: point mutations and module skipping lead to chemical diversity. *Angew. Chem. Int. Ed. Engl.* 45, 2296–2301, doi:10.1002/anie.200503737 (2006). [PubMed: 16506259]
33. Mootz HD et al. Decreasing the ring size of a cyclic nonribosomal peptide antibiotic by in-frame module deletion in the biosynthetic genes. *J. Am. Chem. Soc.* 124, 10980–10981 (2002). [PubMed: 12224936]
34. Gao L et al. Module and individual domain deletions of NRPS to produce plipastatin derivatives in *Bacillus subtilis*. *Microb. Cell Fact.* 17, 84, doi:10.1186/s12934-018-0929-4 (2018). [PubMed: 29855381]
35. Yu D, Xu F, Zhang S & Zhan J Decoding and reprogramming fungal iterative nonribosomal peptide synthetases. *Nat. Commun.* 8, 15349, doi:10.1038/ncomms15349 (2017). [PubMed: 28534477]
36. Linne U & Marahiel MA Control of directionality in nonribosomal peptide synthesis: role of the condensation domain in preventing misinitiation and timing of epimerization. *Biochemistry* 39, 10439–10447 (2000). [PubMed: 10956034]
37. Belshaw PJ, Walsh CT & Stachelhaus T Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science* 284, 486–489 (1999). [PubMed: 10205056]
38. Lu L et al. Mechanism of action of thalassospiramides, a new class of calpain inhibitors. *Sci. Rep.* 5, 8783, doi:10.1038/srep08783 (2015). [PubMed: 25740631]
39. Fischbach MA & Clardy J One pathway, many products. *Nat. Chem. Biol.* 3, 353–355, doi:10.1038/nchembio0707-353 (2007). [PubMed: 17576415]
40. Firm RD & Jones CG Natural products—a simple model to explain chemical diversity. *Nat. Prod. Rep.* 20, 382–391 (2003). [PubMed: 12964834]
41. Davidsen JM & Townsend CA In vivo characterization of nonribosomal peptide synthetases NocA and NocB in the biosynthesis of nocardicin A. *Chem. Biol.* 19, 297–306, doi:10.1016/j.chembiol.2011.10.020 (2012). [PubMed: 22365611]
42. Thirlway J et al. Introduction of a non-natural amino acid into a nonribosomal peptide antibiotic by modification of adenylation domain specificity. *Angew. Chem. Int. Ed. Engl.* 51, 7181–7184, doi:10.1002/anie.201202043 (2012). [PubMed: 22711659]
43. Zhang JJ, Yamanaka K, Tang X & Moore BS Direct cloning and heterologous expression of natural product biosynthetic gene clusters by transformation-associated recombination. *Methods Enzymol.* 621, 87–110, doi:10.1016/bs.mie.2019.02.026 (2019). [PubMed: 31128791]
44. Zhang H, Fang L, Osburne MS & Pfeifer BA The Continuing Development of *E. coli* as a Heterologous Host for Complex Natural Product Biosynthesis. *Methods Mol. Biol.* 1401, 121–134, doi:10.1007/978-1-4939-3375-4\_8 (2016). [PubMed: 26831705]
45. Datsenko KA & Wanner BL One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U S A* 97, 6640–6645, doi:10.1073/pnas.120163297 (2000). [PubMed: 10829079]
46. Tang X et al. Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. *ACS Chem. Biol.* 10, 2841–2849, doi:10.1021/acscchembio.5b00658 (2015). [PubMed: 26458099]
47. Choi KH & Schweizer HP mini-Tn7 insertion in bacteria with single attTn7 sites: example *Pseudomonas aeruginosa*. *Nat. Protoc.* 1, 153–161, doi:10.1038/nprot.2006.24 (2006). [PubMed: 17406227]
48. Davison J et al. Insights into the function of trans-acyl transferase polyketide synthases from the SAXS structure of a complete module. *Chemical Science* 5, 3081–3095, doi:10.1039/c3sc53511h (2014).

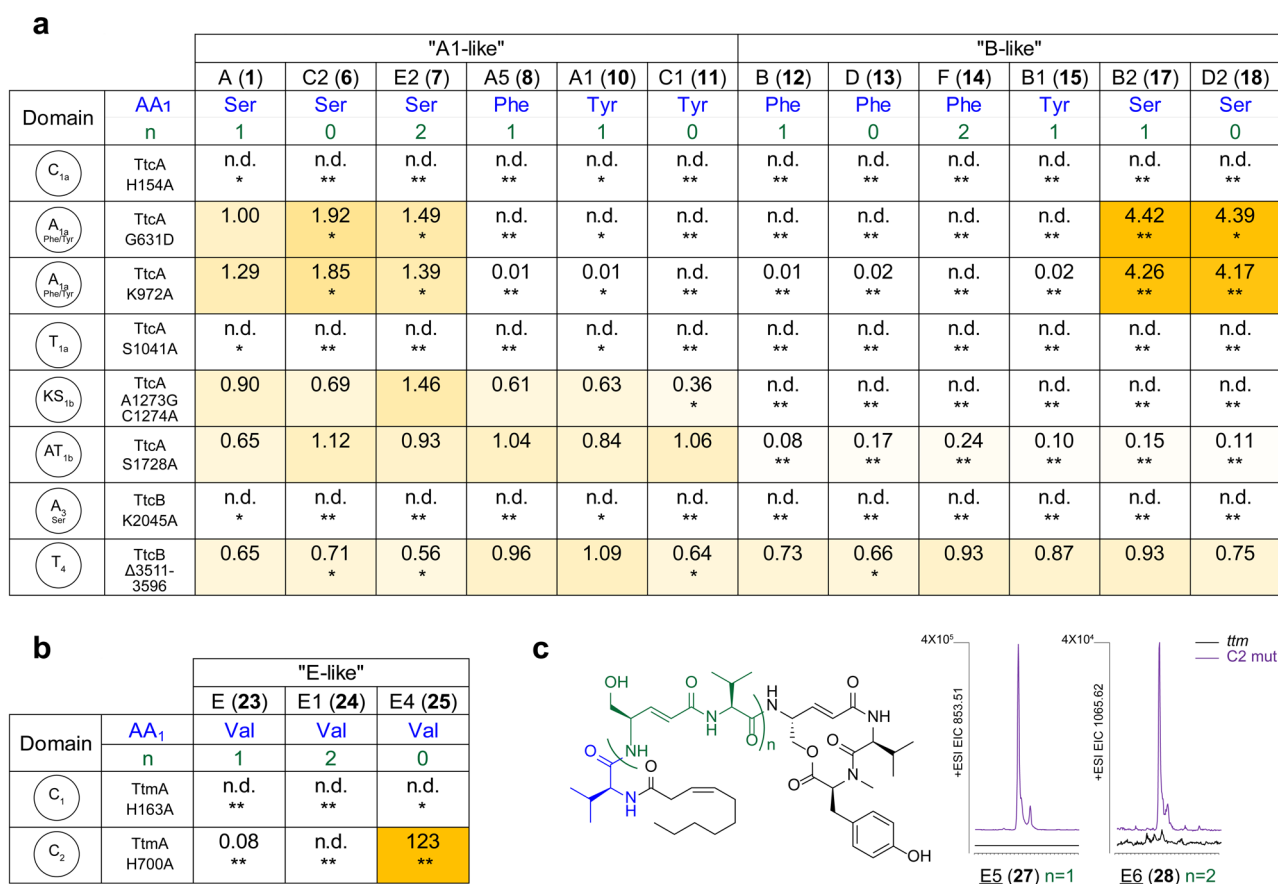
49. Tanovic A, Samel SA, Essen LO & Marahiel MA Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science* 321, 659–663, doi:10.1126/science.1159850 (2008). [PubMed: 18583577]
50. Dutta S et al. Structure of a modular polyketide synthase. *Nature* 510, 512–517, doi:10.1038/nature13423 (2014). [PubMed: 24965652]
51. Edwards AL, Matsui T, Weiss TM & Khosla C Architectures of whole-module and bimodular proteins from the 6-deoxyerythronolide B synthase. *J. Mol. Biol.* 426, 2229–2245, doi:10.1016/j.jmb.2014.03.015 (2014). [PubMed: 24704088]
52. Keatinge-Clay A Crystal structure of the erythromycin polyketide synthase dehydratase. *J. Mol. Biol.* 384, 941–953, doi:10.1016/j.jmb.2008.09.084 (2008). [PubMed: 18952099]
53. Akey DL et al. Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway. *Structure* 18, 94–105, doi:10.1016/j.str.2009.10.018 (2010). [PubMed: 20152156]
54. Al-Mestarihi AH et al. Adenylation and S-methylation of cysteine by the bifunctional enzyme TioN in thiocoraline biosynthesis. *J. Am. Chem. Soc.* 136, 17350–17354, doi:10.1021/ja510489j (2014). [PubMed: 25409494]
55. Miller BR, Sundlov JA, Drake EJ, Makin TA & Gulick AM Analysis of the linker region joining the adenylation and carrier protein domains of the modular nonribosomal peptide synthetases. *Proteins* 82, 2691–2702, doi:10.1002/prot.24635 (2014). [PubMed: 24975514]
56. Jiang W, Bikard D, Cox D, Zhang F & Marraffini LA RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* 31, 233–239, doi:10.1038/nbt.2508 (2013). [PubMed: 23360965]
57. Sawitzke JA et al. Recombineering: highly efficient in vivo genetic engineering using single-strand oligos. *Methods Enzymol.* 533, 157–177, doi:10.1016/B978-0-12-420067-8.00010-6 (2013). [PubMed: 24182922]



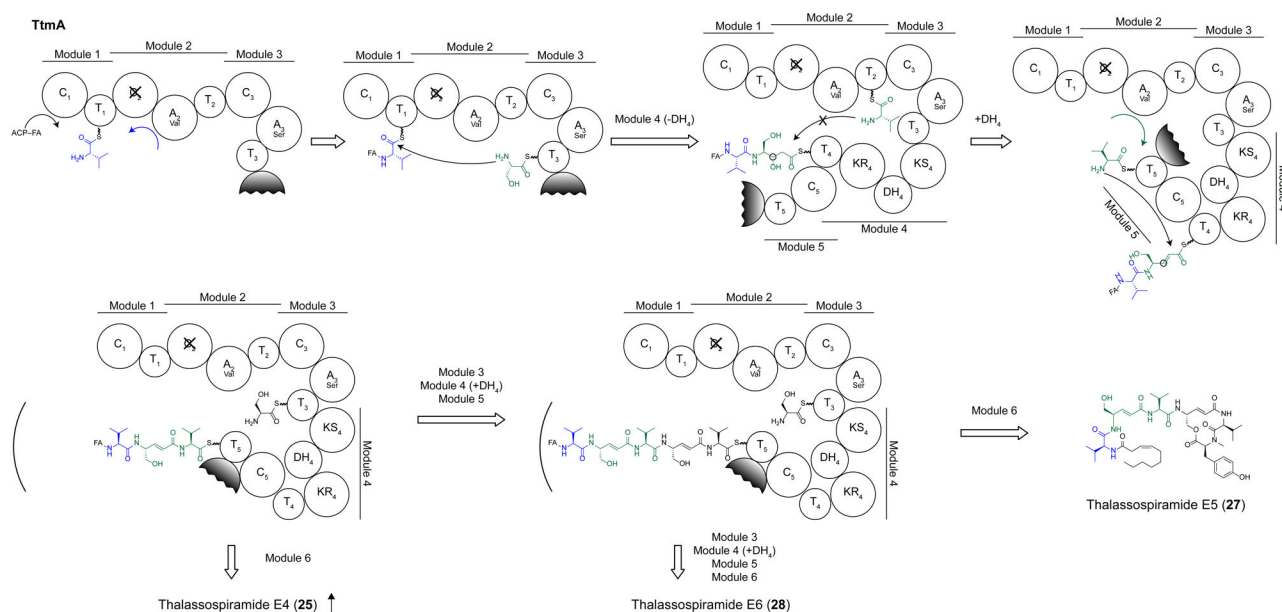
**Fig. 1 |. Heterologous reconstitution of thalassospiramide biosynthetic gene clusters in a *P. putida* host.**

**a,b,** Annotated genomic loci encompassing the thalassospiramide assembly line genes from *Thalassospira* sp. CNJ-328 (**a**) and *Tistrella mobilis* KA081020-065 (**b**) targeted for cloning and heterologous expression. (**c**) LC-MS analysis of extracts from an empty *P. putida* EM383 host and hosts with genomically integrated *ttc* and *ttm* pathways compared against an authentic thalassospiramide A (**1**) standard. Extracted ion chromatograms (EIC) of *m/z* 958.5496. See Supplementary Fig. 4 for associated MS/MS spectra.





**Fig. 3 | Selective inactivation of assembly line enzymatic domains alters product formation.**  
**a,b,** Changes in production level of thalassospiramide analogs from Ttc (**a**) and Ttm (**b**) assembly lines upon selective inactivation of specific enzymatic domains. Domain and precise amino acid mutations are listed in the first two columns. Fold-change in MS ion intensity is indicated by number and color intensity, while white boxes indicate analogs were not detected (n.d.) from the mutant. Statistical significance was calculated using a two-tailed Student's t-test; n=3 biologically independent samples, \*p<0.05, \*\*p<0.005. (**c**) Structures and EICs of new thalassospiramide analogs produced upon TtmA C<sub>2</sub> inactivation; see Supplementary Table 2 for HR-MS data.



**Fig. 4 I. Model for thalassospiramide biosynthesis by Ttm C<sub>2</sub> inactivation mutant.**

Valine is loaded onto T<sub>1</sub> by A<sub>2</sub> and C<sub>1</sub> catalyzes addition of an activated fatty acid (FA) bound to an acyl carrier protein (ACP) or coenzyme A. Module 2 is skipped, and the fatty valine is extended directly onto serine-loaded T<sub>3</sub> via C<sub>3</sub>. Chain extension proceeds normally to module 4, where the substrate is not immediately dehydrated and would normally be passed back to module 2 via C<sub>2</sub>, perhaps based on chain length. However, C<sub>2</sub> inactivation forces the intermediate forward, where it becomes transiently accessible to DH<sub>4</sub> and is dehydrated before extension onto T<sub>5</sub>, which is loaded with valine by A<sub>2</sub>. Direct progression through module 6 results in formation of thalassospiramide E4 (**25**), which is substantially increased as a result of C<sub>2</sub> inactivation. Alternatively, passage from module 5 back to module 3 results in generation of new thalassospiramide analogs E5 (**27**) and E6 (**28**), which undergo one or two additional rounds of chain extension, respectively, through modules 3-5.



Serine is loaded onto T<sub>1a</sub> by A<sub>3</sub>, and C<sub>1a</sub> catalyzes addition of an activated fatty acid (FA) bound to an ACP or coenzyme A. Module 1b is skipped, and the fatty serine is passed directly to valine-loaded T<sub>2</sub> via C<sub>2</sub>. Chain extension proceeds normally to module 4, where the substrate is sequestered from DH<sub>4</sub>, perhaps within an enzyme binding pocket. The intermediate is passed from module 4 back to module 2 via C<sub>2</sub>, where it undergoes another round of chain extension through modules 3 and 4. Upon return to module 4, the intermediate is now acted upon by DH<sub>4</sub>, perhaps due to a conformational change associated with longer chain length, and extended forward through modules 5 and 6 to generate thalassospiramide A (**1**).